Centrality maps and the analysis of city street networks

Thomas Courtat
Orange Labs
38-40 rue du Général Leclerc
92794 Issy-les-Moulineaux, France
thomas.courtat@orange-ftgroup.com

Stéphane Douady
Laboratoire de Matière et Systèmes Complexes (MSC)
UMR CNRS- Université Paris Diderot CC 7056
10 rue Alice Domon et Léonie Duquet
75205 Paris Cedex 13, France
stephane.douady@paris7.jussieu.fr

Catherine Gloaguen
Orange Labs
38-40 rue du Général Leclerc
92794 Issy-les-Moulineaux, France
catherine.gloaguen@orange-ftgroup.com

ABSTRACT

Firstly introduced in social science, the notion of centrality has spread to the whole complex network science. A centrality is a measure that quantifies whether an element of a network is well served or not, easy to reach, necessary to cross. This article focuses on cities' street network (seen as a communication network). We redefine two classical centralities (the closeness and the straightness) and introduce the notion of simplest centrality. To this we introduce a mathematical framework which allows considering a city as a geometrical continuum rather than a plain topological graph. The color plotting of the various centralities permits a visual analysis of the city and to diagnose local malfunctionings. The relevance of our framework and centralities is discussed from visual analysis of French towns and from computational complexity.

Keywords

Centrality, Simplest distance, Street networks, City modelling

1. INTRODUCTION

A city's street network forms a transportation, exchange and communication system. The scientific literature has been interested in cities since Auerbach and Zipf formulated laws for the distribution of cities and populations sizes [3, 22]. Numerous economical considerations of the city as a thermodynamic system have followed [6]. Then the formation of the city overall shape has been modelled using various methods: cellular automata [9, 2], fractals [14, 2], DLA [19] and the differentiation of space by means of multi-agent systems [2, 21]. But most of these models work on square grids or consider units of build-up areas in an Euclidian space.

Only lately with the advent of complex network theory [8] physicists took the street system into consideration: its diversity [10], its topology [10, 16, 20, 11] and its formation [5]. By studying theoretical random walks on streets, [7] has shown that the shape of the network determines in a preponderant way the behavior of agents on the network. When considering a city's street system as a network (or large graph with non trivial features) the analysis has focused on the characterization of the topology [16, 10] and of the efficiency [11] for which the notion of centrality [13, 18, 20] is an important point. A centrality is a measure defined whether on the vertices or the edges of a graph that quantifies to what extent that element is an important link in the whole system. All these studies are made on small square extractions of maps [16, 20] and consider the city as a purely topological object, i.e it is a graph whose vertices are intersection between streets and edges are portions of streets.

In fact a city's resources do not gather on intersections but are distributed all along the streets [17]. The purpose of this article is to construct a mathematical framework allowing to model a city as a spatial graph whose each point in the edges are also vertices. In section 2 we recall the notion of geometrical and straight graphs defined in [12]. The graphs are provided with a measure that integrates information all along the edges. We reconstruct one of the most important features of the city by introducing a hypergraph structure that defines streets as an alignment of street segments (edges).
Sectin 3 redefines in this context two classical centralities (closeness and betweeness previously stuided in a discrete form in [18, 13, 20]) and introduces the notion of simplest centrality.
Finally, section 4 discusses the relevance of our framework and of the centralities by their usefulness in map analysis and their computer-time complexity.
Our data are the vector extraction of a whole city's street network from a Geographical Information System.

2. A MATHEMATICAL FRAMEWORK TO HANDLE WITH CITY MAPS

We restrict a city to its map: the network of its streets without additional information such as population density, traffic or space use.
In this context, cities have been represented by planar graphs [20, 7] whose edges are portions of streets and vertices their intersections. Nonetheless this topological representation

VALUETOOLS 2011, May 16-20, Paris, France

DOI 10.4108/icst.valuetools.2011.245740

has some drawbacks. At first it strongly distinguishes intersections that would be points of interest and streets that would simply bind them. In fact, a city is a distributed system and each point in the geometry of the edges should be considered with the same importance (think of shops distributed all along the streets and not only at their intersections). Secondly, vertices of degree two are taken into account in this graph representation; what sense do they make? They are not intersections but additional points used by the map-maker to sample curved streets into straight segments. Our purpose here is to define a space to represent the idea of "continuous graph": a mathematical object that features a topological graph skeleton and an homogenous distribution of resources on edges.

2.1 Geometrical and straight graphs

A graph $G = (V, E)$ is a finite number of vertices V and a part E of $V \times V$. If $\sharp V$ is large one would prefer to use the world network. If V are points in an Euclidian space we speak of spatial networks [4] and if elements of E are materialized by geometrical curves that intersect only at their extremities that are elements of V we say here that we are in touch with a geometrical graph. Hence a geometrical graph is both a topological object (from (V, E)) and a geometrical one (elements of E are curves). When elements of E are segments, we say G is a straight graph.

For a straight graph, $V = (v_1, ..., v_n)$ and the adjacency matrix $A = (a_{ij})$ totally define the graph. For city maps A is a sparse matrix so we will prefer to represent each element of E with a reference to its extremities in V and calculate A when it is necessary (3.1, 4.3).

2.2 Hypergraph structure

A hypergraph is a graph whose edges can contain several vertices. One can transform a graph $G = (V, E)$ into a hypergraph by using an equivalence relationship on E. The idea here is that for a city map, edges are "street segments" and that we can reconstruct the notion of street (a chain of aligned street segments) with an equivalence relationship. Let (V, E) a graph and R a reflexive relationship on E^2. Then the relationship $\hat{R}$ (transitive closure) defined by :

$$e_1 \, \hat{R} \, e_2 \text{ iif } \exists \, \alpha_1 = e_1, \, \alpha_2, \, \ldots, \, \alpha_n = e_2 \in E \, | \\ \alpha_1 \; R \; \alpha_2, \, \alpha_2 \; R \; \alpha_3, ..., \alpha_{n-1} \; R \; \alpha_n \qquad (1)$$

is an equivalence relationship. From this, one can consider R_θ :

$$e_1 \; R_\theta \; e_2 \quad \text{iif} \quad (e_1 \star_2 e_2) \vee ((e_1 \star e_2) \wedge (|\measuredangle(e_1, e_2) - \pi| \leq \theta)) \qquad (2)$$

where $e_1 \star e_2$ means that e_1 and e_2 intersect, and $e_1 \star_2 e_2$ that e_1 and e_2 intersect in a vertex of degree 2. $\measuredangle(e_1, e_2)$ stands for the geometric angle between e_1 and e_2.
This R_θ allows recovering the notion of "streets" even if input data do not contain such labels. The algorithm labeling streets segments with a street number does not depend on its starting point and is fast to run ($0(\sharp V)$ with optimized structures). The price to pay is that some special cases as forks of two segments making a very small angle with a third one will be considered as a single street.
Fig. 1 shows a straight graph and its additional hypergraph structure. Hypergraphs are represented by a graph (vertices and edges) and by a vector with as much elements as edges. The value of the vector's i-th component is the number of the street to which the i-th edge belongs. Fig. 2 is the pseudo-code of a recursive algorithm to construct an hypergraph from a graph and a relationship.

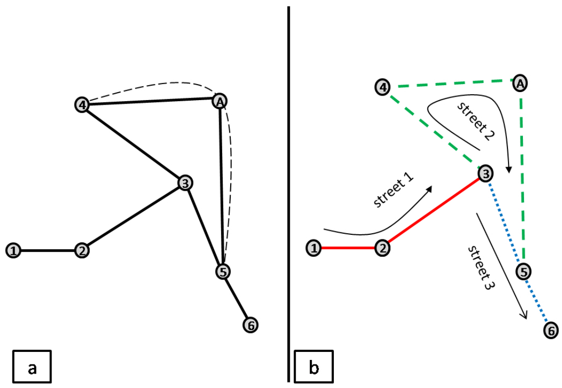

Figure 1: A straight graph (a) and its hypergraph structure (b) deduced from $R_{\pi/20}$. Viewed as a city's map, this graph contains 7 streets segments but 3 streets. The dotted line is an actual curved street from 4 to 5, sampled into straight segments by adding the point A.

2.3 Measure and Monte-Carlo estimation

To a geometrical graph G one can associate its geometrical projection:

$$\pi_G = \{x \in \mathbb{R}^2, \exists e \in E, x \in e\} \qquad (3)$$

π_G will be in the following identified to G. G is a compact space and thus we define its borelian σ-algebra and its borelian measure μ_G. For instance $\mu_G(G)$ is the total length of edges in G. Or

$$\Psi^*(\alpha) = \mu_C \left(c \in C, \measuredangle(c, \vec{u}_0) \in [0, \alpha]\right) \qquad (4)$$

is the total length of the city that makes an angle in $[0, \alpha]$ with a reference vector $\vec{u}_0$ and thus $d\Psi^*(\alpha)/d\alpha$ is the angular density of the graph. If f is a measurable function on G, we seek out to estimate $\int_G f(g) d\mu_G(g)$. An easy to implement algorithm is a Monte-Carlo method. The measure $P_G(.) = \mu_G(.)/\mu_G(G)$ is a "uniform" probability measure on G with its borelian σ-algebra. A random variable X following P_G provides a random point on the graph G. In addition,

$$\mathbb{E}(f(X)) = \frac{1}{\mu_G(G)} \int_G f(g) d\mu_G(g) \qquad (5)$$

and if $X_1, ..., X_n$ are n independent random variables that follow P_G,

$$\frac{\mu_G(G)}{n} \sum f(X_i) \rightarrow \int_G f(g) d\mu_G \qquad (6)$$

which allows to approximate any integral on G. In practice, to get a random point on a straight graph with respect to P_G, one chooses at first an edge with a probability proportional to its length. We then pick a random point on this edge, that is seen as a barycenter of the two extremities.

```
function hypergraph H = graphToHypergraph(Graph graph)

        e= card(graph.edges)
        H= zeros(1, e)
        cursor =1   % the number of the current edge
        currentNum =1  % the number of the current street we are labeling

        while(cursor <= e)
                if( H(cursor) = 0)  % if the edge has not been labeled yet
                        H = treat(H, graph, cursor, currentNum)
                        % label the edge and edges in the same street
                                currentNum = currentNum +1
                                % go to next street
                end
                cursor = cursor +1  % go to the next not labeled edge
        end
end function

sub boolean b = relationship(Edge edge1, Edge  edge2)
% b is TRUE iif edge1 R edge2

sub rec Array H2 = treat(Array H, Graph  graph, Int cursor, Int currentNum)

        H2= H % copy the data
        H2(cursor)=currentNum % label the edge
        N= neighbours(graph, graph.edges(cursor))
        % find all the edge that intersect with the cursor-th edge
        n= card(N)

        for(i= 1 to n)
                if(relationship(N(i),  graph.edges(cursor))
                % if the relationship is true
                        H2 = treat(H2, Graph, N(i), currentNum)
                        % we treat recursively the neighbour
                end
        end
end
```

Figure 2: The algorithm adding a hypergraph structure to a graph. H is an vector with the same number of elements as graph.edges. At the end $H(i)$ contains the number of the street to which the i-th edge belongs. The main function *graphToHypergraph* recursively calls the sub function *treat* that labels an edge.

2.4 City graph

A city graph G is a straight graph that represents a city street network. Most of the time actual streets are straight. If it is not the case, it is possible to sample a geometrical graph to a straight graph and keeping the planarity (Fig 1). We provide G with an hypergraph structure: $G = ((V, E), H)$. Elements in E are called street segment and those in H streets. $V = V_1 \cup V_2 \cup V_+$ where V_1 contains degree 1 vertices called dead-ends, V_+ vertices of degree > 2 called intersection and V_2 vertices of degree 2 (junctions) seen as sampling artifacts to simplify intricate curves into straight segments. For a vertex$v$ we write $E(v)$ the set of edges that pass through v, for an edge e, $V(e)$ the extremities of e and for any point g in G, $E(g)$ the street segment on which lies g and $V(g) = V(E(g))$.

3. CENTRALITIES

3.1 Closeness and straightness

There are two natural distances on a geometrical graph: the Euclidian d^e and the shortest path one d^{sp}. In the complex network framework [1, 8] a centrality is a measure defined on V that quantifies whether a node has a central location or not. Among existing centralities (closeness, betweeness, straightness, information), we focus here on those that directly derive from distances.

The closeness centrality C_i^C of a node i measures if the node is close to the others in average.

$$C_i^C = \frac{\sharp V - 1}{\sum_{j \in V \ , \ j \neq i} d^{sp}(i, j)} \tag{7}$$

The straightness centrality C_i^S of a node i measures if the node is rather in a straigth line or not and if that node efficiently transforms euclidian distances into shortest path distance

$$C_i^S = \frac{1}{\sharp V - 1} \sum_{j \in V \ , \ j \neq i} \frac{d^e(i, j)}{d^{sp}(i, j)} \tag{8}$$

These two centralities can be redefined by means of μ_G for each point x of the city graph:

$$C^C(x) = \frac{\mu_G(G)}{\int_G d^{sp}(x, g) d\mu_G(g)} \tag{9}$$

$$C^S(x) = \frac{1}{\mu_G(G)} \int_G \frac{d^e(x, g)}{d^{sp}(x, g)} d\mu_G(g) \tag{10}$$

$C^S(x)$ is well defined since $d^e(x, g) \underset{x \to g}{\sim} d^{sp}(x, g)$. The shortest path distance can be computed easily from the Floyd-Warshall algorithm [15] that provides shortest-path distances between all pair of vertices and for x, y two points in G : $d_C^{sp}(x, y) = \min_{e_x \in E(x) \ , \ e_y \in E(y)} \{d_C^{sp}(e_x, e_y) + ||e_x - x|| + ||e_y - y||\}$.

3.2 The simplest distance

A plausible behavior for a human being to go from place A to place B would be to adopt *the simplest path* instead of the shortest one. To model this choice, we define the information distance $d_{C,p}^i(A, B)$ from a point to another along a path p in the city. And we define the information distance between those two points as : $\min_{p \text{ path from } A \text{ to } B} d_{C,p}^i(A, B)$.

Let's compute this distance from the toy map Fig. 3. To direct somebody from A to B in along the doted red path (P_1) instructions are: (1) From A take the street to the right with respect to the house, (2) go straight (keep on the same street) while meeting 6 street-crossings, (3) on the 7th intersection, take the second street from the right, (4) go straight - 4 streets-crossings, (5) on the 5th intersection take the third one, (6) go straight - 1 streets-crossing, (7) walk for 35 meters. (an intersection of degree n counts for n streets-crossing).

This is encoded in the so called "path information vector" $\vec{I}(A \to B||P_1) = [1, 6, 2, 4, 3, 1, 35]$ and for P_2 the blue doted path that is the shortest path between A and B, $\vec{I}(A \to B||P_2) = [1, 0, 2, 0, 1, 1, 2, 0, 2, 0, 2, 0, 115]$. The information length of apath is its number of components: 7 and 13 respectively for P_1 and P_2. We define the information distance between two points as the minimal information distance among all paths that go from A to B. In Fig. 3 the red dotted path was the simplest so the information distance from A to B is 7.

By convention, the simplest distance between A and B $d^{sim}(A, B)$ is 0 if $A = B$, 1 if A and B are in the

same street and the information distance minus 2 otherwise.

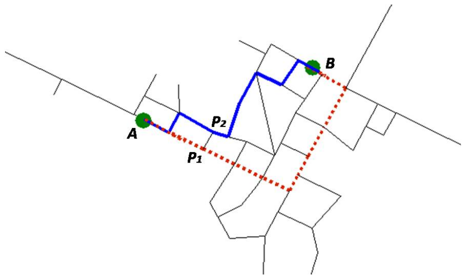

Figure 3: Two points A and B located on a city graph. The blue solid path is the shortest path between the two points and the red dotted one is the simplest path.

3.3 Simplest centrality

As for the closeness, the averaging of the simplest distance defines a centrality (the simplest centrality):

$$C^{sim}(x) = \frac{\mu_G(G)}{\int_G d^{sim}(x,g)d\mu_G} \quad (11)$$

Since C^{sim} is constant on a street h. Thus the centrality can be defined on H.
Let's define the street adjacency matrix H^*, $H^*(h_i, h_j) = 1$ if h_i and h_j intersect and 0 otherwise.

The topological distance is the shortest path distance calculated on H^*. It counts the number of times one needs to turn to go from a street to another one. We can calculate all topological distances between pairs of streets by calculating a street adjacency matrix H^* with $h_{ij} = 1$ if i-th and j-th street intersect and by applying a Floyd algorithm.

The simplest centrality rewrites:

$$C^{sim}(x \in h_0) = \frac{\mu_G(G)}{\sum_{h \in H} ||h||.d^{top}(h, h_0)} \quad (12)$$

A street that maximizes C^{sim} is called a center of the city h_c. If the street network is not too regular h_c is unique. The topological radius of the network is : $r^{top} = \max d^{top}(h, h_c)$ and the diameter is:

$$\text{diam}^{top} = \max_{h_1,h_2 \in H} d^{top}_C(h_1, h_2) \quad (13)$$

with $\text{diam}^{top} \leq 2.r^{top}$

4. DISCUSSION

4.1 Visualization

A centrality is a function from a map to $\mathbb{R}$ so three dimensions are needed to visualize it. We represent the street network in $\mathbb{R}^2$ and associate to each point a color that cods for its centrality. To this we perform a mapping from the image of the centrality to a color map with C colors. The mapping is not linear: we used a histogram equalization in order to get a more contrasted plot. That way a color represents a proportion $1/C$ of the city and the color steps in every $1/C$-quintile.

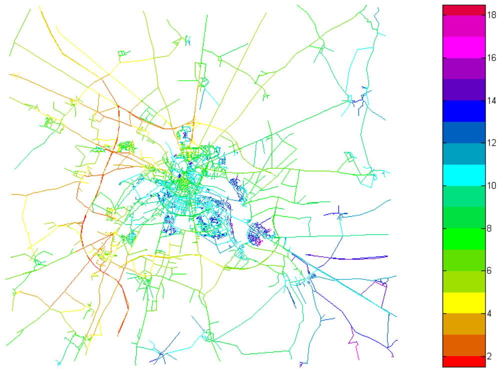

Figure 4: The topological distance to the center for the city of Amiens and its nearest suburb. The center is the highway belt and the rest of the city organizes hierarchically: there is no radial component in the map, a few long streets maintain a topological distance almost constant in every zone.

Each centrality produces an interpretation of the map: a radial one for the closeness, a local one for the straightness and a hierarchical one for the simplest centrality. This is illustrated in Fig. 4, 5.

For the closeness, the main effect is a side effect: if for Troyes the maximal centrality coresponds to the city center, for Avignon the maximum is hit in the center of the image rather than in the center of the city (on the top left). This leads to think that in "normal" cities, $d^e \simeq d^{sp}$. The straightness is side effects free but the overall impression is ill-assorted. The simplest centrality provides a hierarchical view of the city. On each map, a zone is particularly striking (middle right) for it produces a discontinuity in the overall variation of the centralities. The centralities are able to put in light ill-deserved zones. Troyes is quite homogeneous but the centrality analysis sticks to reality in Avignon: the main axes are the old city walls, a high way and radial axis. The ill-deserved zone detected on the right is a residential area, volontary isolated and the one on the bottom is poor area known by town planners.

In Figure 4 we seek for the simplest center of the extended city of Amiens and map each street to its simplest distance to that center. The center happens to be the highway belt, which is not surprising since it is indeed its purpose to reduce distances in a city. Then a few long streets grid the whole area and consequently the variation of the distance is not radial but constant in the overall with local hierarchical variations. Besides the radius of $r^{top} = 18$ is small in comparison to the size of the city.

4.2 The importance of geometrical graphs

Representing cities by straight graphs rather than plain graphs, using μ_G, is important wor several reasons. Firstly it allows a continious plot of the centrality rather than a discrete one as in [13]. Secondly, when calculating a centrality on a plain graph one introduces a bias due to vertices of degree 2. Indeed in the data this kind of vertices is artificially added

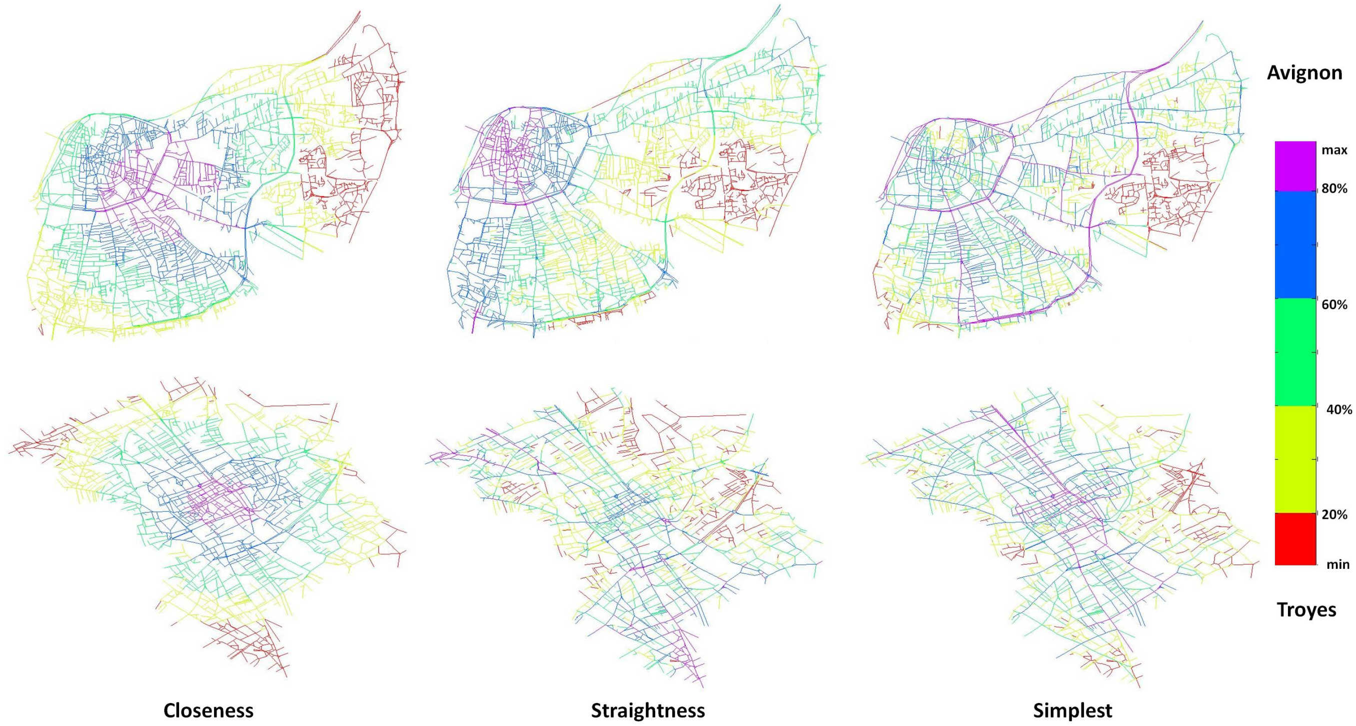

Figure 5: (Color online). The different centralities studied in this article for two french cities: Avignon (the extended city center) and Troyes. The color mapping is obtained with 5 colors and a histogram equalization so that in each picture a color represent a proportion $1/C$ of the city.

to sample curved roads to straight segments. Thus curved streets have a bigger weight than straight ones. Moreover the proportion of vertices of degree two is quite variable according to the city: 0.1 in Troyes, 0.14 in the center of Amiens and 0.34 in the whole Amiens (Fig 4).

4.3 Complexity

The computation is faster for the simplest centrality than for the closeness or the straightness ones. We start with the skeleton of a city graph G onto the form of two arrays V for vertices and E for edges. V is of size $v \times 2$ and contains coordinates of vertices. E is of size $e \times 2$ and contains the references in V of the edge end points. Let's note $h = \sharp H$ and $L = \mu_G(G)$

The steps of the calculation of the straightness or the closeness are:

1. Compute the adjacency matrix: $0(e)$
2. Compute the skeleton of shortest path distances with a Floyd algorithm (three nested loops): $\alpha.v^3$.
3. Estimate the centrality on samp.L points each with samp.L points (sampling): $O(\text{ech}^2.L^2)$

As for the simplest centrality:

1. Compute the hypergraph: $O(e^2)$
2. Compute the street adjacency matrix: $O(ev)$
3. Compute the shortest path: $\alpha.h^3$
4. For each street calculate the centrality: $O(h^2)$ (with a very small coefficient).

In average all v, e, h, L are linked together and we consider that $e \simeq 1.5v$, $h \simeq 0.25e \simeq 0.37v$. Hence theoretically the simplest centrality calculus is about 20 times faster than the other ones'.

5. CONCLUSION

In this article we have presented a mathematical framework (geometrical graphs) that allows to consider a city map not as a purely topological object but as a continuum. In addition, we have shown that a relationship based on local alignment of street segments permits to reconstruct the notion of street. We have redefined in this context two classical centralities (closeness and straightness) that derive from Euclidian and shortest-path distances as well as a new one: the simplest centrality that can be seen as the integration of the simplest distance. Each of these centralities allow to locate malfunction in a city transportation network. The closeness is above all radial with a strong side effect and the straightness is a local measure whereas the simplest centrality offers a hierarchical view of the city, main roads as highway belts, former surrounding walls. The calculation of the simplest centrality is faster than the other ones.

6. REFERENCES

[1] R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of modern physics*, 74:47–97, 2002.

[2] S. Albeverio, D. Andrey, P. Giordano, and A. Vancheri. *The Dynamics of Complex Urban Systems*. Springer, 2008.

[3] F. Auerbach. Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59:74–76, 1913.

[4] M. Barthélemy. Spatial networks. *arXiv:1010.0302v2*, 2010.

[5] M. Barthélemy and A. Flammini. Co-evolution of density and topology in a simple model of city formation. *Networks and spatial economics*, 9:401–425, 2009.

[6] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West. Growth, innovation, scaling, and the pace of life in cities. *PNAS*, 104:7301–7306, 2007.

[7] P. Blanchard and D. Volchenkov. *Mathematical analysis of urban spatial networks*. Springer Complexity, 2009.

[8] S. Boccalettia, V. Latorab, Y. Morenod, M. Chavezf, and D.-U. Hwanga. Complex networks: Structure and dynamics. *Physics Reports*, 424:175–308, 2006.

[9] H. L. Bras. *Essai de géométrie sociale*. Editions Odile Jacob, 2000.

[10] J. Buhl, J. Gautrais, N. Reeves, R. Solé, S. Valverde, P. Kuntz, and G. Theraulaz. Topological patterns in street networks of self-organized urban settlements. *Eur. Phys. J. B*, 49:513–522, 2006.

[11] A. Cardillo, S. Scellato, V. Latora, and S. Porta. Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73:066107, 2006.

[12] T. Courtat, C. Gloaguen, and S. Douady. Mathematics and morphogenesis of the city: a geometrical approach. *Physical Review E*, 2010.

[13] P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, 73:036125, 2006.

[14] P. Frankhauser. The fractal approach. a new tool for the spatial analysis of urban agglomerations. *Population*, 10:205–240, 1998.

[15] J. L. Gross and J. Yellen. *Handbook of graph theory*. CRC Press, 2003.

[16] B. Jiang and C. Claramunt. Topological analysis of urban street networks. *Environment and Planning B : Planning and design*, 31:151–162, 2004.

[17] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore. Scale invariance in road networks. *Physical Review E*, 73:026130, 2006.

[18] V. Latora and M. Marchiori. A measure of centrality based on the network efficiency. *arXiv:cond-mat/0402050v1*.

[19] H. A. Makse, J. S. Andrade, and M. Batty. Modeling urban growth patterns with correlated percolation. *Physical Review E*, 58:7054–7066, 1998.

[20] S. Porta, P. Crucitti, and V. Latora. The network analysis of urban streets: a primal approach. *Environment and Planning B: planning and design*, 33:705–725, 2006.

[21] F. Sembolonia, J. Assfalgb, S. Armenib, R. Gianassib, and F. Marsonic. Citydev, an interactive multi-agents urban model on the web. *Computers, Environment and Urban Systems*, 28:45–64, 2004.

[22] G. Zipf. *Human Behaviour and the Principle of Least Effort*. Harvard University Press, 1949.